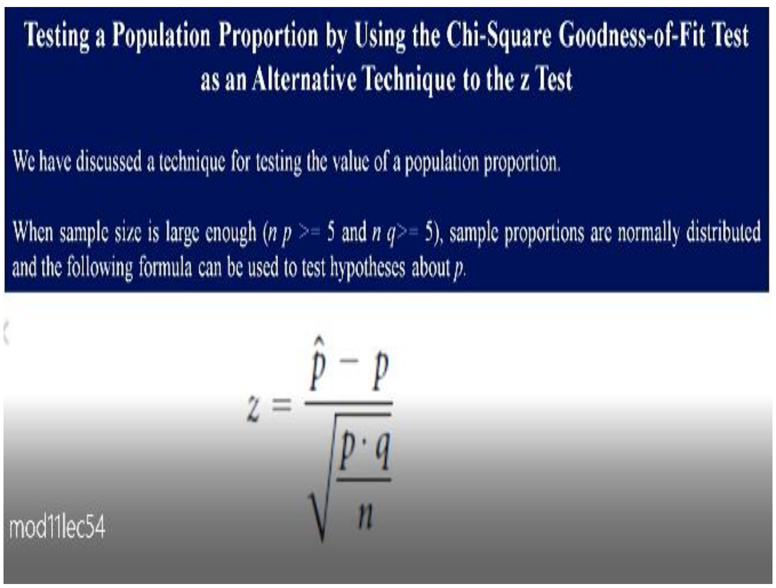
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology - Roorkee

Lecture – 54
Chi-Square Test: Goodness of Fit – II

Hello friends, welcome you all in this session as you are aware in previous session, we were discussing about a technique which we generally use for analysing categorical data or non-numeric data and the technique Chi square goodness of fit wherein in previous class if you look at we discussed whether the expected frequencies and observed frequencies are same or not if they are same then the Chi square value would be 0.

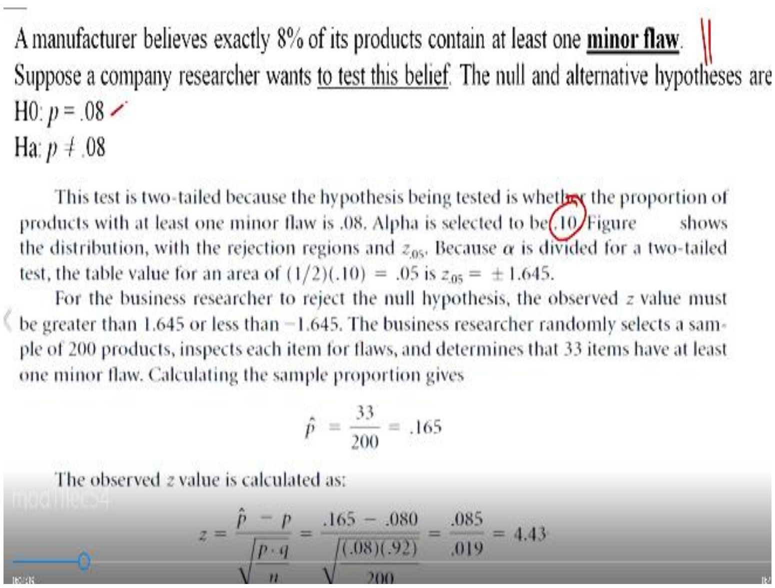
Otherwise, it would be having some positive value let us look at some other applications of Chi square goodness of fit it can be used to as a testing for a proportion hypothesis testing of proportion of 1 sample 2 sample more than 2 samples so let us look at how this technique can be used instead of using Z test okay.

(Refer Slide Time: 01:40)



So let us take a manufacturer in fact before that let me give you a brief theory about this technique wherein we are using a goodness of fit test in place of Z test right. So when sample size is large enough we always say that the sample proportions are normally distributed and the following formula which we have used for calculating value of Z when we solved the questions on hypothesis testing of proportions 1 sample right. So let us look at a question like this.

(Refer Slide Time: 02:18)

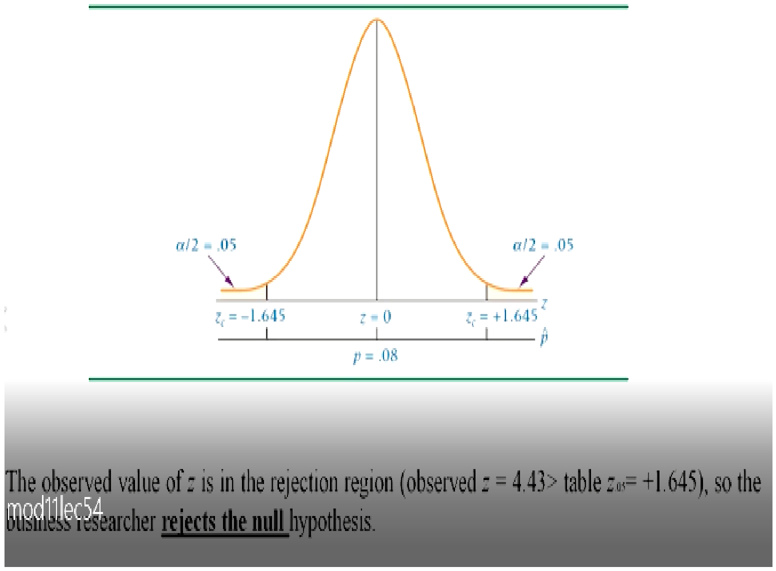


Let us say the engineer at the shop floor things that the 8% of products contains at least one minor flaw and this is the hypothesis which he wants to test at 0.10 significance level right. So null hypothesis is that the manufacturing defects are 8% or 0.08 alternative hypothesis is not =0.08 so it is a case of 2 tailed test right is it not. So what the manufacturer or the manager at this shop floor does is he selects 200 products and he found only 33 items have got minor flaws.

Right so the p value or $\hat{p}$ value is 0.165 and the value of Z is 4.43. Now since the significance level is 95% no its not 95 it is 90%. So the value of Z is 1.64± right so if you if you want to reject the null hypothesis then the Z value should be more than 1.65 or <-1.64 right. So if you look at this Z calculated is this it means when you are rejecting null hypothesis so this -1.64 +1.64 and this is your Z value somewhere here.

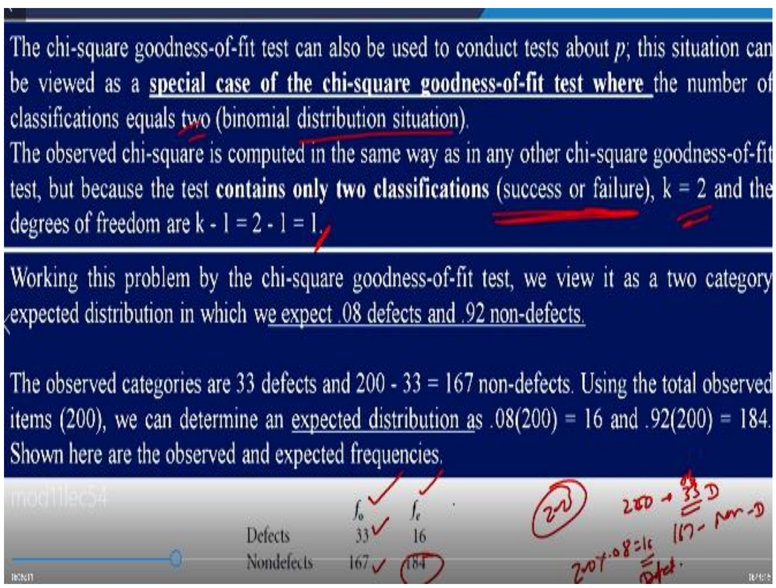
In the rejection region we reject the null hypothesis and we say that no the 8% products do not have minor flaw. Now this is what we have done earlier.

(Refer Slide Time: 04:35)



But the same example can be solved using goodness of fit test okay.

(Refer Slide Time: 04:41)

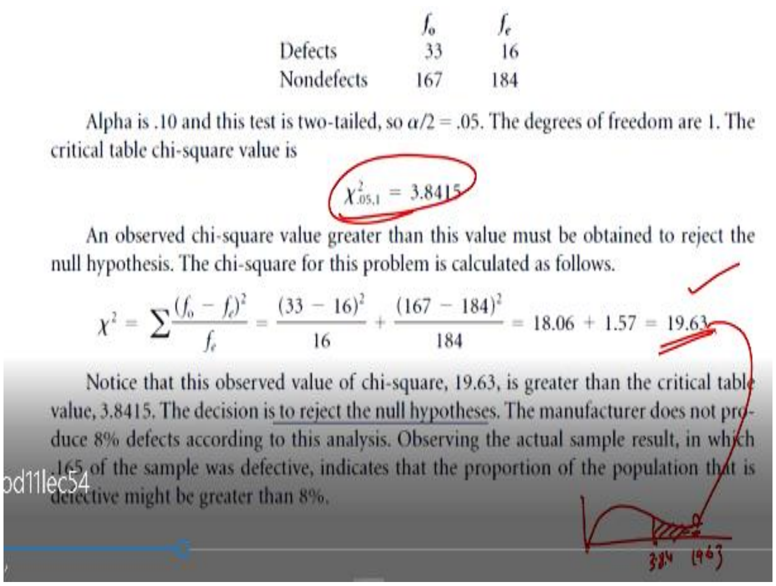


So we will say that the this can be this type of situation can be viewed as a special case of Chi square goodness of fit test where the number of classification =2 right so binomial distribution situation is there right so there are just 2 classifications we just reach are success and failure right. So here you have got k=2 and the degrees of freedom would be k -1 right so this is one degree of freedom right.

And we know that the observed frequency is this out of 200 let us say 200 samples 33 are defective right. So it means 167 are non-defective is it not these are defective these are non-defective. So these are the observed frequencies now just look at this these are the observed frequencies of defectives and non-defectives right. How to get expected frequencies since we know the +200 our assumption about population parameter is 0.08 right.

So just multiply 200 and .08 so it comes out to be 16 so 16 are defective right and I and when 16 are defect to how many are non-defective 184 it is 200-16, 184 so you have got now observed frequencies and the expected frequencies now you can apply Chi square test and you can check whether observed frequencies and expected frequencies match with each other or not right.

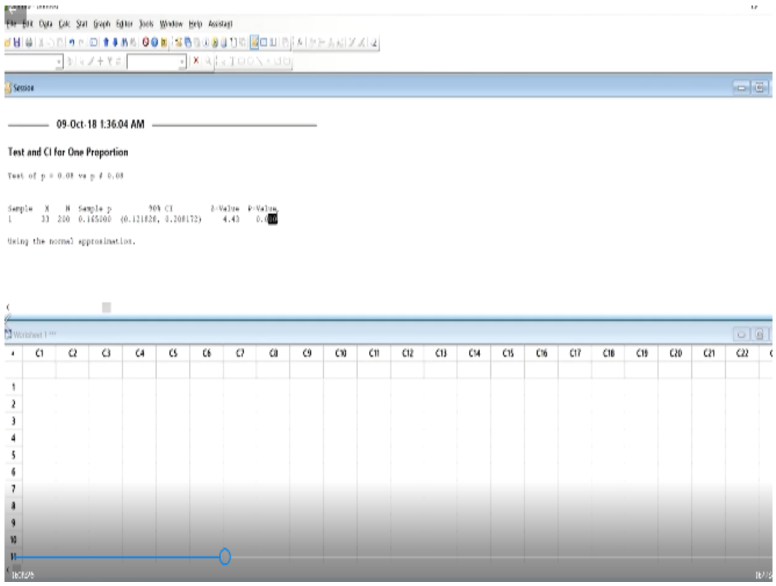
(Refer Slide Time: 06:43)



So the chi square from table is 3.84 and from calculation the Chi square statistics is 19.63 so we will reject the null hypothesis because this is your rejection region right the critical values 3.84 right somewhere here and this is your somewhere here right 19.63 which is in reject so reject the null hypothesis right what was null hypothesis here that observed frequencies and the expected frequencies are same.

So we are rejecting it right and the same and the conclusion here is similar to what we concluded when we solved the question using hypothesis testing of proportion right. So let us look at this particular question and using Minitab right we will solve this question using Minitab.

(Refer Slide Time: 07:36)

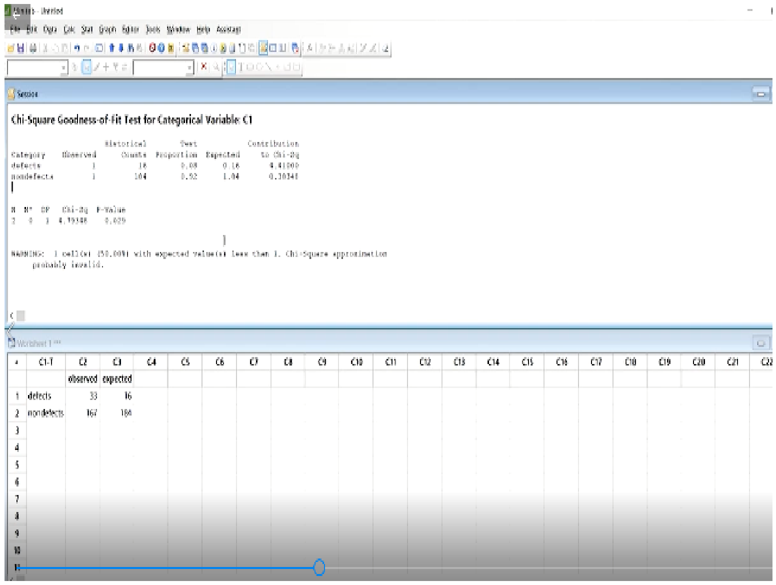


So 1st we will solve this question, one sample proportion test right so 1 P one proportion test. So we need to enter summarized data so number of events is you have got 33 and trials were 200 right so we need to perform hypothesis hypothesize proportion was 0.08 right option at 90% significance level right so method is normal approximation so alternative hypothesis is just keeping my it is not equal to type right so it would be a 2 tailed test right okay click ok.

So p value is this right p value is 0 right so let us compare it with alpha value. So here alpha value is 0.10 right and p is 0 is p < alpha yes so we will reject null hypothesis that is what we have done it right and you can look at other values as well okay you have got this confidence interval yeah Z value 4.43 this is what we calculated right just look at this 4.43 is it not. So just now you have solved this question using Minitab as one sample proportion test right.

You need to solve the same question using Chi square goodness of fit test. So let us look at this question 1st this these are the data points right and this is yeah software right so let us go for data entry 1st.

(Refer Slide Time: 09:36)



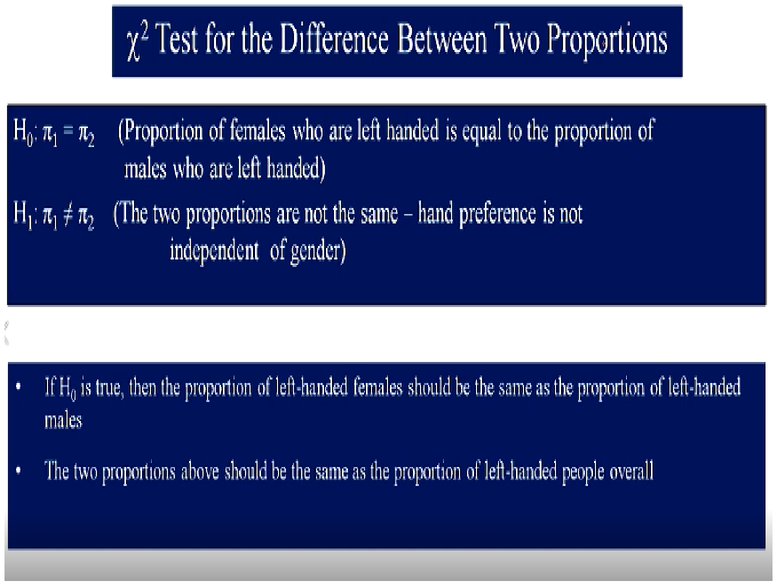
So you have got defects and non-defects right so these are the Fo observed frequencies and the Fe expected frequencies 33 and 167 so the total has to be 208 expected 16 and 184 let us find out what is the value of Chi square good so tables Chi square goodness of fit so observed count this is observed count categorical data. Okay of the proportions specified by historical counts right so expected values need to come over here right.

So okay so let us look at Chi square value yeah so 1st look at the p value was 0.02 right so 0.02 which is p value < alpha no so we will not reject null hypothesis let us look at once again 1st look at Chi square value it is 4.73 let us look at whether are getting the goodness or not so the Chi square calculated is 19.63, 33 167 we will go for data entry 33 167 16 and 184 so if you look at p value is 0.02.

So p value is 0.02 and alpha is 0.01 is p <alpha no so we will not reject null hypothesis according to this it means there is some problem in data entry right. Any way you can in fact we have worked out couple of examples using Chi square goodness of fit test so that you can perform no not a problem anyway we will look at some more examples wherein we will get Chi square value right.

So let us look at next question so what we have done so far is this we have calculated or we have used Chi square goodness of fit test as a replacement of Z test right now the Chi square goodness of fit can also be used for comparing 2 proportions right.

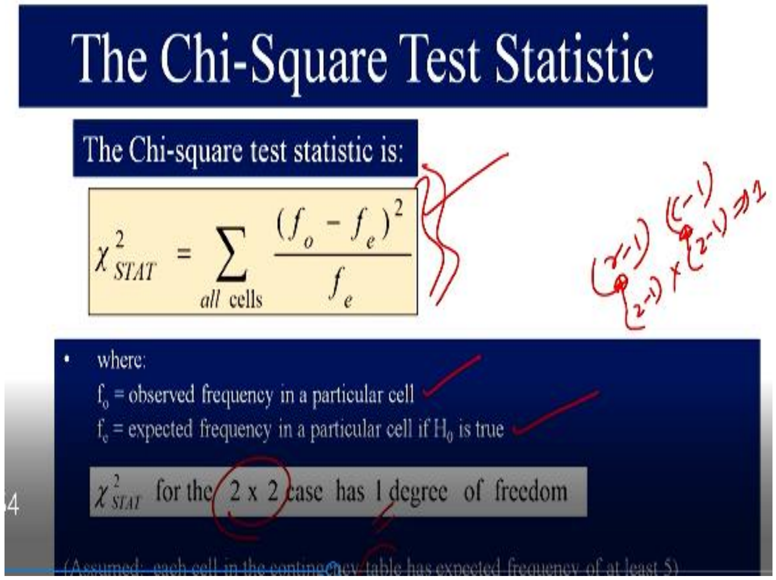
(Refer Slide Time: 13:16)



So you will have π1and π2 let us say the number of absenteeism in a company amongst males and females are same right. So you have got let us say 15% male are absent and 17% females are absent. So we can compare these 2 proportions are same or not okay so let us look at this question just for example. Proportion of females were lefthanded=proportion of males were left handed right so this is your null hypothesis.

And they are not same is alternative hypothesis. If H0 is true, then proportion of lefthanded females should be same as proportion of left handed males right.

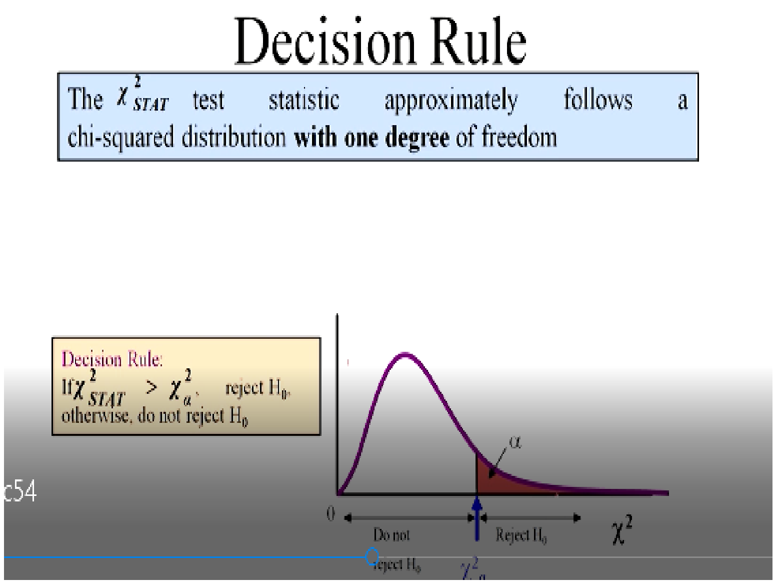
(Refer Slide Time: 14:14)



So how to proceed for comparing 2 proportions. 1st calculate Chi square statistics right so its same as what we have done so far so this is observed frequencies if the expected frequencies Chi square state for 2*2 case for one degree of freedom so how to find out this one degree of freedom it would be like r-1 *c-1 so r is number of rows and c is number of columns. So if there are 2 rows and 2 columns so it would be 2-1 2-1 so it would be just one right.

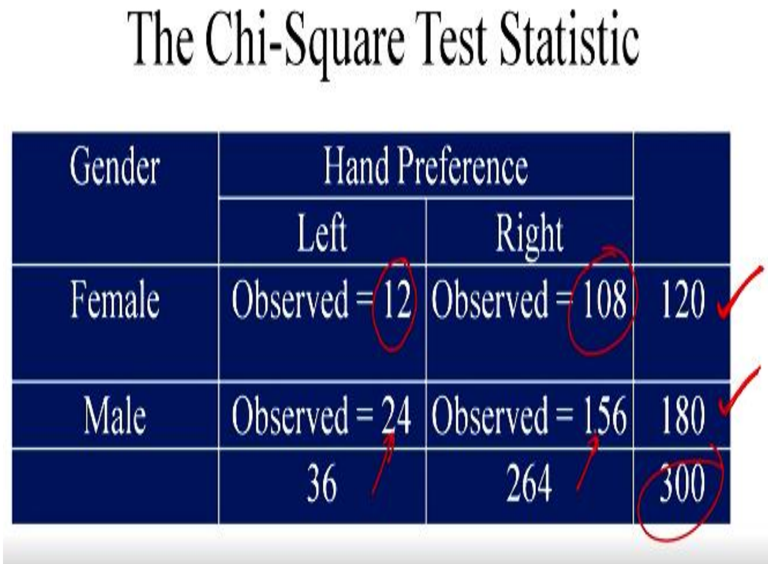
And you should keep in mind that the value in each of the cells of the contingency table should be at least 5 right.

(Refer Slide Time: 15:10)



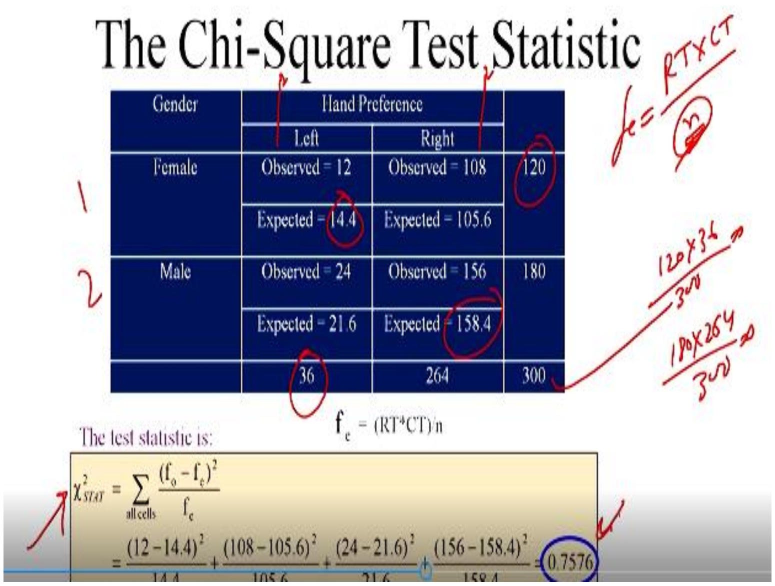
So the decision rule remain same if the Chi square calculated is more than the critical value we will reject the null hypothesis otherwise we will not reject it.

(Refer Slide Time: 15:26)



1st look at this question the Chi square test statistic so let us say you collected data from 300 people out of them 120 were females and 180 were male right so 120 females 180 males. Out of 120 females you have observed 12 were lefthanders 108 were right. Similarly, out of 180 males 24 left lefthanders and 156 right handers right. So you have got these observed frequencies.

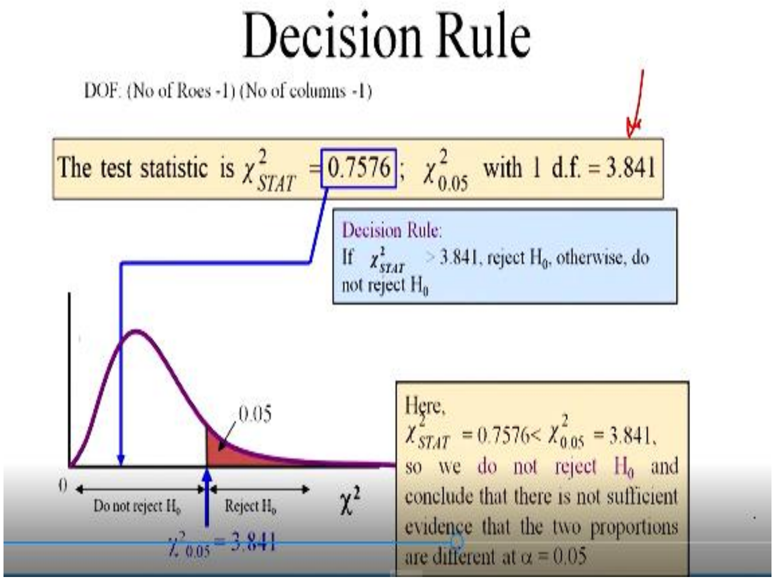
(Refer Slide Time: 16:20)



Now you need to calculate Chi square statistics. So for that 1st of all you need to calculate the expected frequencies is it not. So expected frequency can be calculated by this formula row total*column total/ total number of people whom you observed right n. So here if you look at there are 2 rows 1st row 2nd row and 2 columns this is 1st column and this is 2nd column right.

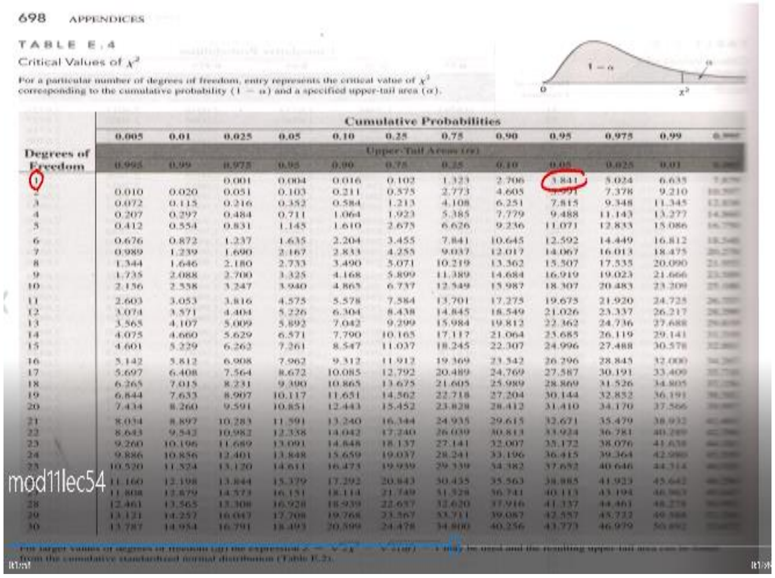
So row total is 120 column total is 36/fe which is 300 you will get expected frequency is 14.4 right. Similarly let us say if you want to calculate expected frequency for this for this particular cell right cell number 22 right 2nd row 2nd column so it would be 180*264/300 okay so expected frequency would be 158.4. Since you have got fo and fe just calculate Chi square great statistics which is 0.75.

(Refer Slide Time: 17:39)



And net appropriate degrees of freedom let us look at this one degree of freedom Chi square value is 3.841 right.

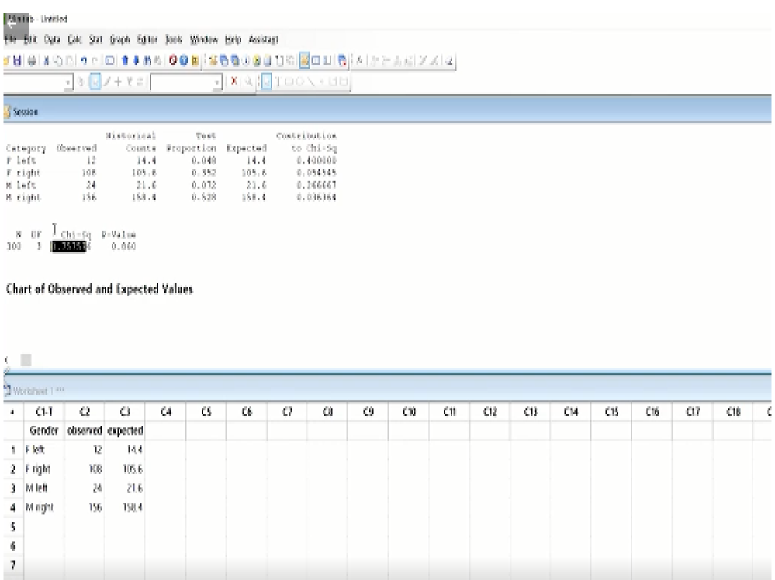
(Refer Slide Time: 17:51)



Let us look at how did we get this 3.814 right so at one degree of freedom right it is 3.841 sorry. Okay this is the critical value 3.841 and the calculated value is 0.7 right the critical value is 3.8 yeah 3.841 and this is calculated value which is in non-rejection yet so we will not reject null hypothesis and well say that the proportion of male and female using left hand are same there is no significant difference right.

So there is not sufficient evidence that the 2 proportions are different okay so let us look at how to solve this question using Minitab. So we will have data like this.

(Refer Slide Time: 18:56)

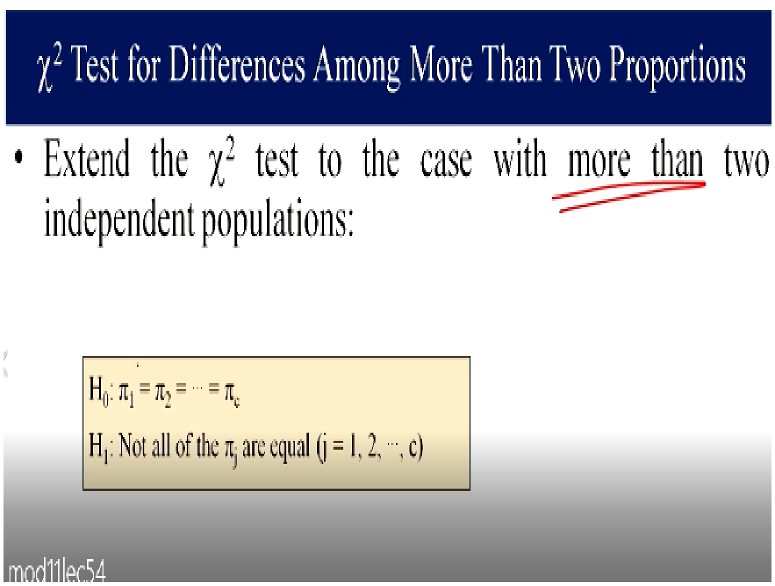


So let us look at this question so you can yeah you should delete all these output 1st so you have gender female left handers and female right hander male left and male right so observed 120 yeah it is 12 108 24 156 14.4 105.6 21.6 158.4 so let us look at Chi square credit goodness of fit so just look at output just look good this value 1st of all p value right compare p with alpha so alpha here was 0.05 alpha 0.05, p value is 0.86 is p value <alpha no so we will not reject null hypothesis.

That is the conclusion we have made earlier as well right now let us look at Chi square statistics which we have calculated and the values 0.7575 right in fact when we did solve previous question we were not getting Chi square because we made some error while entering data right. But this the way he knows you should enter data so this is how you can calculate the 2 proportions.

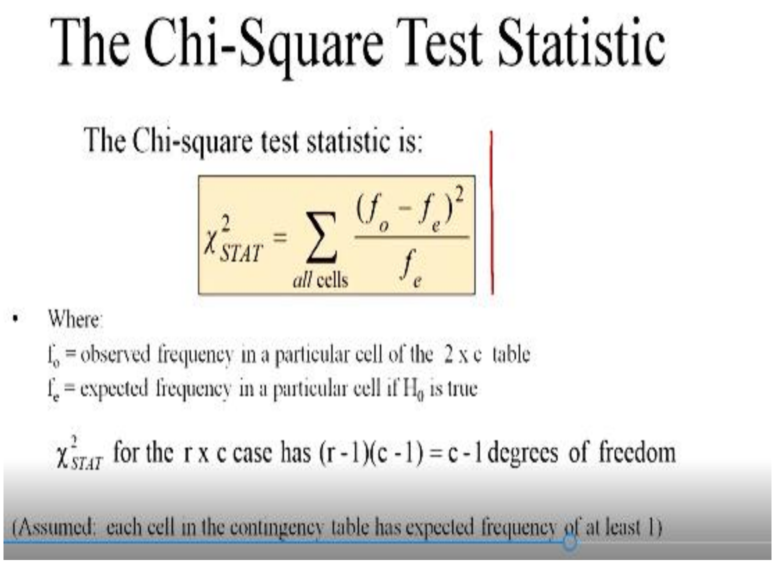
And you can find out whether there is a significant difference exist or not right. Now what if the proportions are if let us say there are 3 groups and you want to compare proportions of those 3 groups let us say the percentage of MBA boys passing a course MBA girls passing a course and PHD scholars passing a course if you want to compare these 3 proportions. So what you should do again you can apply.

(Refer Slide Time: 22:23)



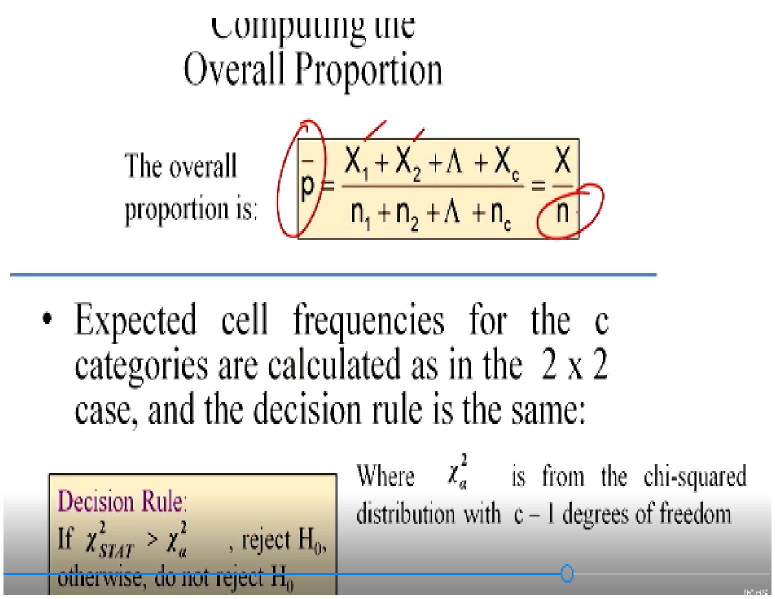
Chi square test of goodness right so you will say that a null hypothesis is this all these proportions are same and they are not the same as alternative hypothesis.

(Refer Slide Time: 22:38)



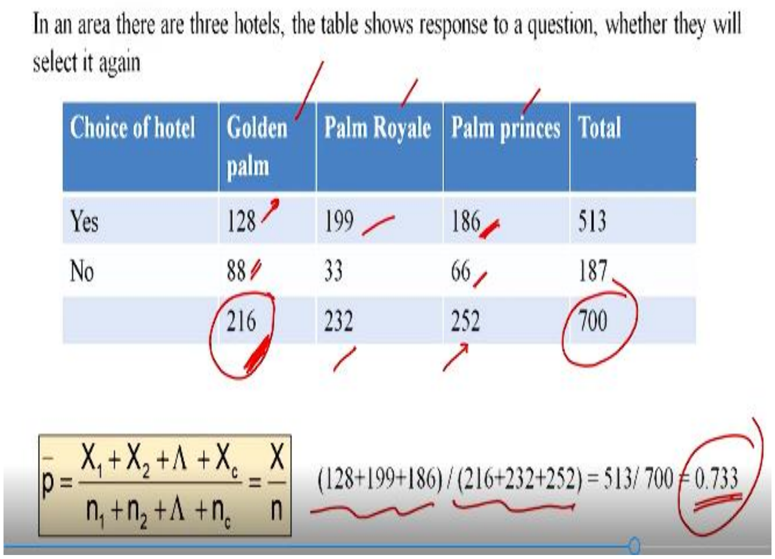
This is same Chi square statistics formula will remain same so observed frequency in a particular cell of 2*c so we are saying that there are 2 rows and c number of columns right. fe expected frequency so the Chi square statistics at appropriate degrees of freedom would be calculated by r-1*c-1 right and r is number of rows and c is number of columns right.

(Refer Slide Time: 23:11)



So 1st of all you should compute the overall proportions when you have got a question like this so this very simple p bar=x1 x2 to xc and the total observations and the decision rule remains same if Chi square statistics calculated is more than critical value we reject the null hypothesis otherwise we will not reject null hypothesis right.

(Refer Slide Time: 23:52)



In an area there are three hotels, the table shows response to a question, whether they will select it again

Choice of hotel	Golden palm	Palm Royale	Palm princes	Total
Yes	128	199	186	513
No	88	33	66	187
	216	232	252	700

$$\bar{p} = \frac{X_1 + X_2 + \Lambda + X_c}{n_1 + n_2 + \Lambda + n_c} = \frac{X}{n}$$

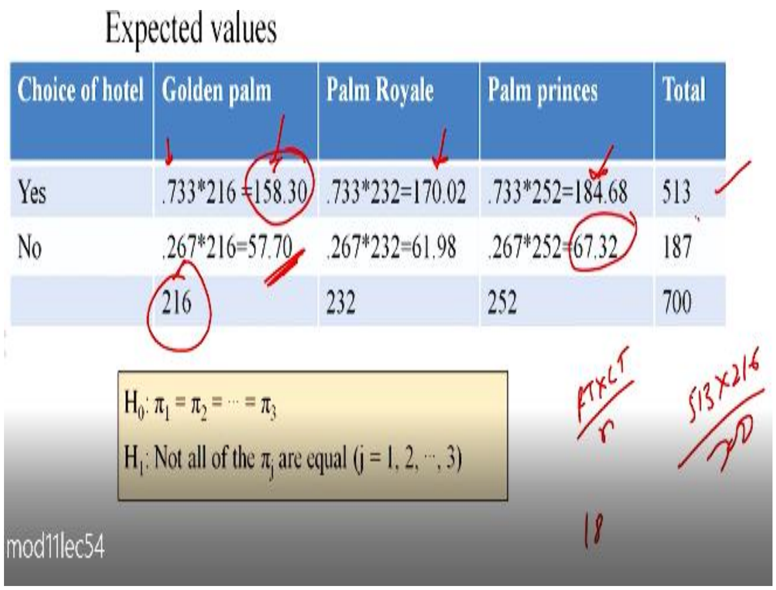
(128+199+186) / (216+232+252) = 513/ 700 = 0.733

Let us look at a question like this let us say in a city or in an area there are 3 hotels right. So there are 3 hotels the table shows response to a question whether they will select the hotel again so there are 3 hotels Golden Palm, Palm Royale and Palm Princess a question was asked to 700 people total right. So when a question was asked to 216 guests of the hotel this hotel and 128 said they would come again in that hotel and 88 said no.

Similarly, in palm princess hotel 186 said they will come again and 66 said no they will not come again. So you need to calculate 1st is called the overall proportion p bar is 0.773 it is very simple. So you have got 128 199 186 /total number of samples selected from each of these hotels right so 216 232 252 216 in other words total said yes/ total number of yes right. So these many percentage of let us say 73.33% of people said yes.

In other words, 1-0.733 said no they will not come again right. So let us find out expected frequencies these are observed frequencies right.

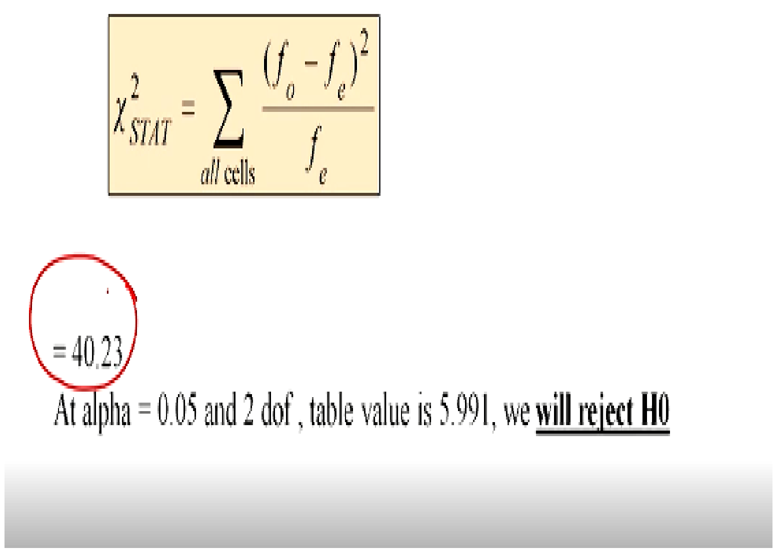
(Refer Slide Time: 25:59)



So for expected frequencies you just multiply these emphasize with proportion for yes right. So it is 0.773*216 so this is the expected frequency similarly this is the expected this is expected frequency this is expected frequency how you get this expected frequency it is 1-0.77 is 0.267 right multiplied by 216. So this is one way of calculating expected frequencies the other method I have already taught you it is row total*column total/n right.

So you can do the same thing here as well right. So let us say if you want to calculate this expected frequency right so row total 513 column total 216/700 you will get 158.3 if you want to get this expected value just to calculate it using row total right total column total 252/700 right.

(Refer Slide Time: 27:21)



So after calculating all fo and fe you need to calculate Chi square statistics and in this question the Chi square statistics is 40.23 and at alpha= this 2 degrees of freedom why 2 degrees of freedom number of row- (row-1) right so (row -1)*(column -1) so how many rows there to 2 column were 3 right this 1*2 which is =2. So that is why 2 degrees of freedom so this is how Chi square distribution would look like this the rejection region.

So the calculated value is 40.23 which is in rejection region and critical value is 5.99. So you will reject the null hypothesis right and what was null hypothesis that all these proportions are same is it not this was our null hypothesis that proportion 1 2 3 are same right. So we are rejecting null hypothesis it means the proportions are not same they are different from each other so in in todays class we have discussed about hypothesis testing of proportions of more than 2 samples.

Though we did work old couple of examples on hypothesis testing of proportions using 2 samples and using Z test but in todays class we have used Chi square goodness of fit test for comparing proportions of more than 2 groups and this is nothing but the extension of what we have seen earlier. So you can solve any question let us say even if you say there are 4 groups so even you compare their proportions as well.

So Chi square goodness of fit is quite a robust material which can handle hypothesis testing of proportion of say 1 sample 2 sample 3 and 4 and so on right. So that is the beauty of chi square goodness of fit test in next class, we will discuss some more on Chi square goodness of fit thank you very much.